

Available online at www.sciencedirect.com**ScienceDirect**

Transportation Research Procedia 3 (2014) 932 – 941

**Transportation
Research
Procedia**

www.elsevier.com/locate/procedia

17th Meeting of the EURO Working Group on Transportation, EWGT2014, 2-4 July 2014,
Sevilla, Spain

Flow Rate Effects on Vehicle Speed at Two Way-Two Lane Rural Roads

Riccardo Rossi ^{a,*}, Massimiliano Gastaldi ^a, Federico Pascucci ^a

University of Padova - Department of Civil, Environmental and Architectural Engineering, Via Marzolo 9, Padova 35131, Italy

Abstract

The flow characteristics of a cross-section are identified by time headway and vehicle speed distributions over time. Knowledge of both headway and speed distributions plays a significant role in several fields of traffic flow analysis and simulation. In particular, we refer to operative analysis of road facilities in interrupted and uninterrupted flow conditions.

This work identifies typical vehicle speed probability density functions (pdf's), for two-way two-lane road segments. The scarcity of empirical findings on these distributions in the literature (particularly for two-way two-lane roads) was one of the reasons for the study.

The endogenous traffic parameters considered to affect speed distributions, flow rate and flow composition were taken into account.

Mixture models for vehicle speeds were estimated.

The data used for statistical analysis of headway and speeds came from traffic observations carried out on the rural road network of the province of Verona (north-east Italy).

In the case in question, regardless of traffic conditions, the Skew-T single model turned out to be the one which best fitted the observed phenomena.

The effect of flow rate in both directions (analysed and opposite) on vehicle speed distributions was also analysed. As expected, in all traffic conditions, the curves move to the left and become narrower as the flow rate increases. This effect seems to be less pronounced, where we measure the effect of the opposite flow rate.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the Scientific Committee of EWGT2014

Keywords: two-way two-lane rural roads; vehicle speed distribution

* Corresponding author. Tel.: +39 049 8275563; fax: +39 049 8275577.

E-mail address: riccardo.rossi@unipd.it

1. Introduction

The flow characteristics of a road cross-section are identified by time headway (TH) and vehicle speed (VS) distributions over time. Knowledge of both headway and speed distributions plays a significant role in several fields of traffic flow analysis and simulation. In particular, we refer to operative analysis of road facilities in interrupted and uninterrupted flow conditions. Studies on VS modeling have been published for many years (Gerlough and Huber (1976); Luttinen (1996); Luttinen (2001); Dey et al. (2006); Jun (2010); Zou and Zhang (2011); Zou et al. (2012)).

The research reported here, following previous works (Rossi and Gastaldi (2012); Rossi et al. (2014)), aims at identifying typical VS probability density functions (pdf's) for two-way two-lane road segments. The scarcity of empirical findings on these distributions in the literature (particularly for two-way two-lane roads) was one of the reasons for the study.

Traffic parameters considered to affect VS distributions, flow rate (FR) and flow composition (percentage of heavy vehicles, %HV) in both directions were taken into account.

A new procedure combining TH trend analysis and clustering analysis on TrendLess Samples (TLSs) based on FR measures in both directions was developed to identify fitting samples.

The data used for statistical analysis came from traffic observations carried out on the rural road network of the province of Verona (north-east Italy).

The paper is organized as follows. Section 2 gives a description of the methodological approach and a detailed description of the analysed study case; comments about flow rate effects on vehicles speed are given in section 3. Concluding remarks and future research directions are presented in Section 4.

2. Methodological approach

This work presents a full procedure for fitting VS pdf's. The procedure, which is general and which can also be used for fitting TH models, is shown in the flow chart of Fig. 1. This study refers only to the left-hand part of the flow chart highlighted in gray.

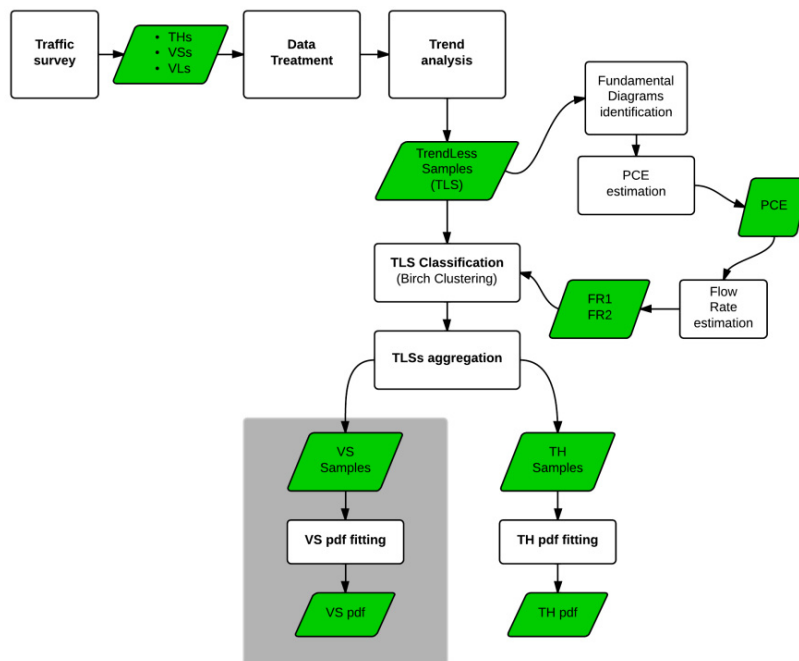


Fig. 1. Main steps of fitting procedure.

A description of each step is briefly presented in the following sections, and refers to the study case analysed here.

2.1. Traffic survey and data treatment

The field data used here came from a survey carried out by the Transportation Laboratory, University of Padova. Two radar sensors, one for direction, were used since it can record THs, together with estimation of vehicle lengths (VLs) and speeds.

Traffic flow observations at a certain point (cross-section) of an entire road segment are only useful for describing its traffic flow characteristics if we accept the hypothesis that the segment is homogeneous (in geometric and functional terms) and that, on the segment in question and over a certain time interval, traffic conditions are steady (constant volume, regardless of the position of the section in the road segment, and time-independent density). In our case, the functional and geometric characteristics were homogeneous along the segment examined (for at least 1 km in both directions).

The traffic data selected came from a cross-section belonging to a road segment with lane widths of 3.60 meters and shoulder widths of 0.60 m; it was located in flat terrain and was perfectly straight. The speed limit was 70 km/h. Traffic data were collected during good weather and dry surface conditions. In this way, exogenous effects (Ha et al., 2011) were assumed to be fixed within the analysed periods.

Table 1 shows the time period and duration of the on-site survey and the main characteristics of traffic data collected for each cross-section examined.

Vehicles longer than 7.5 m were considered to be “heavy”.

Table 1. Duration of traffic survey and main characteristics of traffic data

Section	Period	Duration	Total Vehicles	Heavy Vehicles
VR	2011.12.01-2011.12.06	6 days	47,297	15.2%

The above traffic conditions cover a range of values from 100 to 1,000 vehicles/lane/hour (Fig. 2a). In some peak periods, metastable conditions (transition between stable and unstable conditions) were observed (volumes of nearly 1,000 vehicles/lane/hour). These findings match previous data about this kind of road, where capacity values are usually around 1,200-1,300 vehicles/lane/hour.

The data were treated so that bicycles passing the cross-section were identified as such. Because their riders keep to the right, it was also assumed that they did not affect TH sequences: each TH proceeding and following bicycles was removed from the data set. Two simple criteria were used to identify bicycles: their VL was less than 3.0 m and their speed was under 30 km/h.

The data set without bicycles was used as input for trend analysis (see section 2.2), although all vehicles were taken into account in the FR estimation (see section 2.3).

2.2. Trend analysis

Trend analysis was implemented in R language, following the procedure of Luttinen (1996). Trend analysis is an essential step in studying TH, because the real properties of this variable can only be inspected with stationary data. In order to obtain stationary data sets, the whole sample (Table 1) was submitted to trend analysis by the “exponential ordered scores trend test” of Cox and Lewis (1966), applied by Luttinen (1996) and slightly modified by Rossi et al. (2014).

The traffic conditions observed at the cross-section during the analysed period are clearly described by the Flow Rate/Space Mean Speed diagrams of Fig. 2a: the set of (FR,SMS) pairs was estimated with reference to the TLSs and then represented in plan FR/SMS. As steady flow conditions were assumed, the SMS was estimated as the harmonic average of vehicle speeds recorded at the section during the TLSs.

The distribution of points representing the state of the analysed section shows how the sampled time periods (Table 1) cover the whole domain of the flow rate, ranging from free flow to a value not far from the capacity of the road segments (1,200-1,300 vehicles/lane/hour).

A total of 182 TLSs was used for the following steps of the procedure.

2.3. Flow rate estimation

Each TLS is characterized by a flow rate (FR, vehicles/hour/lane) and by a percentage of Heavy Vehicles (%HV) in both the analysed direction and the opposite one.

In general terms, for the same traffic volume, traffic composition affects traffic operations in significant ways. For this reason, the effect of impedance caused by heavy vehicles cannot be omitted from the analysis. In order to bear this effect in mind, mixed flow must be converted to equivalent passenger-car flow by the passenger-car equivalent (PCE). This equivalent depends mainly on %HV and FR in both directions and should be estimated by taking all these factors into account.

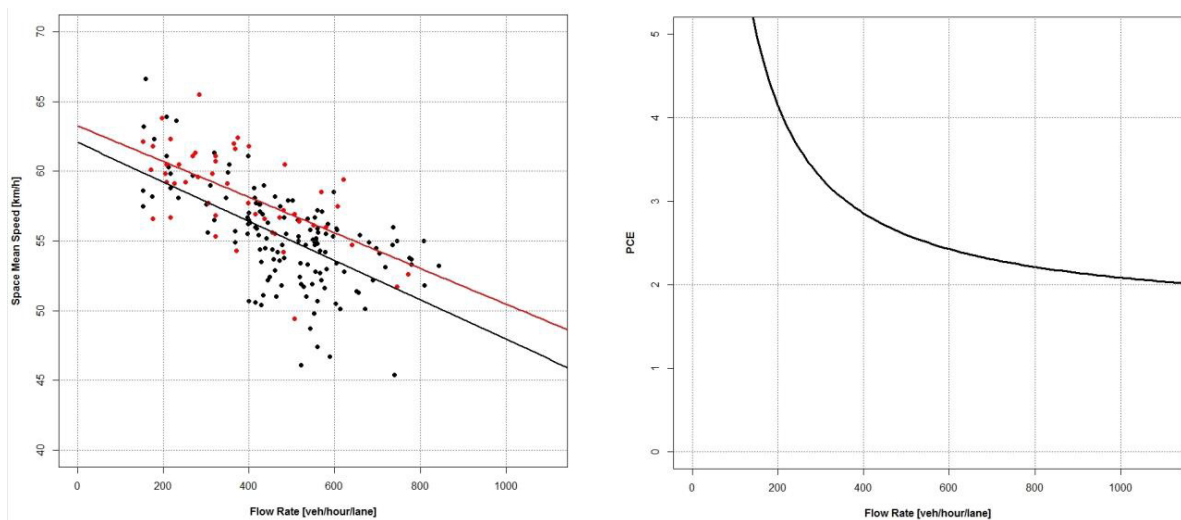


Fig. 2. (a) SMS/FR pairs estimated for each TLSs time interval; (b) PCE as a function of FR in the analysed direction.

In this work, we assumed that PCE depended on FR in the analysed direction, and thus used only one value of the %HV, corresponding to the mean value over the whole set of TLSs ($\overline{\%HV} = 12.6\%$).

In Fig. 2a,

- the red dots represent TLSs with %HV less than or equal to 2% and this condition was assumed to be representative of passenger-car traffic;
- the black dots represent TLSs with %HV over 2%, assumed as representative of mixed traffic. The
- red line shows the linear regression of the red dots (Basic Model, BM), and
- the black line that of the black dots (Mixed Model, MM).

The PCE was computed for each TLS with the formula derived from Luttinen (2001):

$$PCE(FR) = 1 + \frac{SMS_{MM}(FR) - SMS_{BM}(FR)}{S_{BM} \cdot \%HV \cdot FR} \quad (1)$$

where:

SMS_{MM} is the SMS computed from MM as a function of FR

SMS_{BM} is the SMS computed from BM as a function of FR, and

S_{BM} is the slope of BM.

$PCE(FR)$ is shown in Fig. 2b.

In the end, FR is converted into pce/hour/lane by the coefficients computed as described above, which lead to FR_{pce} .

In this way, two FR_{pce} , one for the analysed direction ($FR_{pce,1}$) and one for the opposite ($FR_{pce,2}$) were associated with each TLS. In this calculation the PCE estimated for the analysed direction was also used for the opposite one.

2.4. Trendless samples classification and aggregation

The aim of this task was to aggregate the TLSs into homogeneous groups, with aggregation attributes as traffic characteristics in both directions ($FR_{pce,1}$, $FR_{pce,2}$). Many clustering techniques have been developed for such a problem; in this work we used the Birch clustering technique (Zhang et al. 1996), since it applies an algorithm for hierarchical clustering which does not require much computation time. Other advantages are that it allows users to specify the desired diameter threshold for clusters and can detect outliers.

After the groups had been detected by grouping all the TLSs in the same cluster, aggregate samples TLS_j , were created, so that larger samples of homogeneous data were available for VS pdf fitting.

With reference to the study case, Birch clustering was carried out in R with the “birch” package (Charest et al., 2012). The algorithm detected 9 clusters, of which 3 were discarded because of their low number of their elements. All the characteristics of the six detected clusters are listed in Table 2; results are also shown in Fig. 3.

The six detected TLS_j were therefore representative of the main traffic conditions of the analysed cross-section in the time period under analysis.

Table 2. Main characteristics of detected clusters.

TLS _j	# of TLSs	FR _{pce,1} (pce/hour/lane)				FR _{pce,2} (pce/hour/lane)			
		Mean	Var	Min	Max	Mean	Var	Min	Max
1	31	263	68	169	432	243	60	126	390
2	39	738	67	633	875	572	56	435	670
3	47	544	67	399	649	614	54	538	739
4	10	629	59	486	700	752	42	685	838
5	23	360	63	200	453	433	71	312	539
6	27	555	64	455	663	447	57	325	521

With reference to each aggregate sample TLS_j , Table 3 shows the statistics of VS.

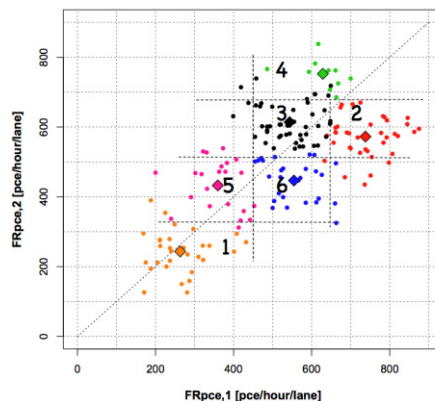


Fig. 3. Detected clusters in plan $FR_{pce,1}$ / $FR_{pce,2}$.

Table 3. VS statistics for each TLS_j (VS in km/h).

TLS _j	Statistics						
	Sample size	Mean	Mode	Median	Var	Skewness	Kurtosis
1	4,097	61,7	60	61	118,7	0,743	1,71
2	11,105	55,2	55	55	76,8	0,1	3,003
3	11,474	56,5	56	56	87,6	0,123	2,193
4	3,130	53	55	53	91,9	-0,202	2,891
5	4,601	60,7	58	60	99,6	0,86	3,312
6	6,620	57,5	56	57	77,7	0,259	3,227

2.5. VS pdf fitting

The aim of this fitting process was to evaluate the theoretical models and to identify the best one in terms of goodness-of-fit.

Three models were chosen to represent VS distribution: Normal (N), Skew-Normal (SN) and Skew-T (ST). The first was chosen because of its ample use in the literature, and SN and ST for their capability to model excess skewness and kurtosis.

In order to capture the multi-modality of speed distributions, a mixture modeling approach was applied; in the analysed case up to three components were introduced into the models. Models characteristics are listed in Table 4.

Table 4. Number of free parameters of analysed VS pdf models as a function of number of components *g*.

Model	Parameters	g=1	g=2	g=3
N	ξ, σ^2	2	5	8
SN	ξ, σ^2, λ	3	7	11
ST	$\xi, \sigma^2, \lambda, \nu$	4	8	12

Estimation of the model parameters for each cluster was carried out by the Expectation-Maximization algorithm (EM) of Dempster et al. (1977), which is an iterative method for finding ML estimates. In this paper, estimation was carried out in R with the “mixsmsn” package, which executes the process with the EM method (Prates et al., 2013).

In the case of mixture models, an initial K-means algorithm is performed to optimize the initial sub-clusters. Because this algorithm uses random initial values of the parameters of its objective function, the fitting process may produce different model parameter estimates in different runs. However, the differences are not substantial: as the number of runs increases, the K-S distance appears to be limited to a narrow range. In the analysed case, 20 runs for each TLS_j were carried out and the optimal solution (value of model parameters) with respect to the K-S test was chosen.

The goodness-of-fit of the calibrated models was evaluated by three indicators:

- the K-S distance revealed the adherence/harmony of the model and was used as a comparison measure among the models;
- Pearson’s chi-squared test was applied to evaluate the acceptability of the model;
- the Bayesian Information Criterion (BIC) was used to identify the most suitable model for all traffic conditions.

The K-S test is widely used to compare a sample with reference probability distribution. It is based on estimation of test statistic *D*, i.e., the absolute vertical distance between the Empirical Distribution Function (EDF) and the fitted model; the test statistic is then compared with some critical values for determining model acceptability (level of significance). This approach has a limitation, in that it can only be applied to continuous distributions: this is a problem when the variables, as in this case, are integers without decimals (because of equipment precision) making the empirical distribution more similar to a discrete than a continuous distribution. In addition, as noted by Zou and

Zhang (2011), “in a K-S test with a sufficiently large sample, a small and non notable statistic D can be found to be statistically significant”. This comment means that we could not use the K-S approach as a test for the acceptability of a model, but only as a useful metric in model comparisons.

For this reason, the criterion for rejecting a model was assumed to be Pearson’s chi squared test. Bin width was set at 1 km/h, corresponding to the precision of VS data.

BIC was used to compare various models and, for the same model, the effect of number of components. BIC can indicate whether the increase in the likelihood of the model due to an increase in the number of model parameters can lead to overfitting. It is based on the sum of two terms, the double-negative log-likelihood value and a term related to the complexity of the model and depending on both sample size and number of free parameters.

The results for the K-S statistic may be summarized as follows:

- for all the TLSj (traffic conditions) and regardless of number of components, ST is the best model: for the sake of brevity, Fig. 4a shows only the comparison among one-component models;
- for all the TLSj, the one-component models (single, $g=1$) behave worse than the mixture ones ($g=2, 3$); in addition, for $g \geq 2$, as the number of components increases the K-S statistic marginally decreases. Fig. 4b shows the effect of number of components on goodness-of-fit with reference to Normal models.

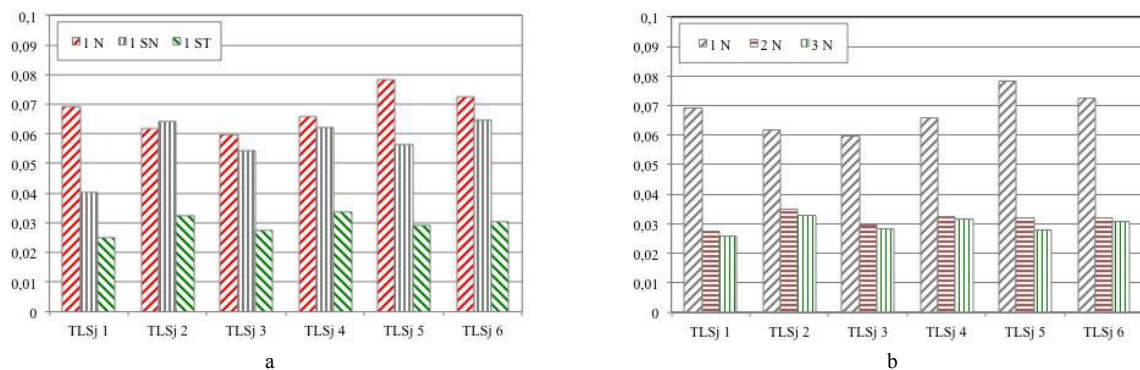


Fig. 4. K-S statistics of six TLSj, comparisons of: (a) single models; (b) Normal models.

With reference to Pearson’s chi-squared test, the results may be summarized as follows (see Table 5):

- N single models were not significant in any traffic condition;
- SN single models were not significant in most traffic conditions;
- all the mixture models are better than the corresponding single ones.

Table 5. Pearson’s chi-squared test. Models p-values for the six TLSj.

	N			SN			ST		
	g=1	g=2	g=3	g=1	g=2	g=3	g=1	g=2	g=3
TLSj 1	0.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00
TLSj 2	0.00	0.99	1.00	0.00	0.67	0.99	1.00	1.00	1.00
TLSj 3	0.00	1.00	1.00	0.00	0.94	1.00	1.00	1.00	1.00
TLSj 4	0.01	1.00	1.00	0.01	1.00	1.00	1.00	1.00	1.00
TLSj 5	0.00	1.00	1.00	0.05	1.00	1.00	1.00	1.00	1.00
TLSj 6	0.00	0.98	1.00	0.00	0.86	1.00	1.00	1.00	1.00

The results for the BIC values are summarized as follows (see Table 6):

- for the same model specification, the number of free components often leads to worsening of the goodness-of-fit of the models;
- for N and SN, single models always behave worse than the corresponding mixture ones;

- ST single models are the best, regardless of traffic conditions.

Table 6. Bayesian Information Criterion. Models BIC values for the six TLS_j.

Model				Model				Model			
g	N	SN	ST	g	N	SN	ST	g	N	SN	ST
1	31,211.7	30,966.9	30,901.1	1	79,739.1	79,714.8	78,721.2	1	83,893.8	83,853.0	83,073.0
2	30,933.1	30,947.3	30,937.9	2	78,741.5	78,841.2	78,743.5	2	83,085.4	83,147.8	83,098.1
3	30,935.9	30,964.2	30,962.9	3	78,739.9	78,783.9	78,780.4	3	83,089.8	83,130.2	83,135.5
TLS _j 1				TLS _j 2				TLS _j 3			
g	N	SN	ST	g	N	SN	ST	g	N	SN	ST
1	23,047.7	23,029.5	22,760.4	1	34,241.0	33,984.4	33,723.8	1	47,620.5	47,554.4	46,957.9
2	22,769.5	22,802.8	22,790.2	2	33,761.8	33,772.9	33,756.1	2	46,989.3	47,056.3	46,989.2
3	22,779.0	22,802.6	22,820.9	3	33,751.5	33,776.1	33,789.0	3	46,969.1	47,004.7	47,011.5
TLS _j 4				TLS _j 5				TLS _j 6			

The set of estimated goodness-of-fit statistics gives a clear-cut result: in all traffic conditions (TLS_j) Skew-T single models represent the best trade-off between goodness-of-fit and model complexity.

The capacity of these models to represent the observed phenomena properly is also confirmed by the comparison between empirical distributions and fitted models (Fig. 5).

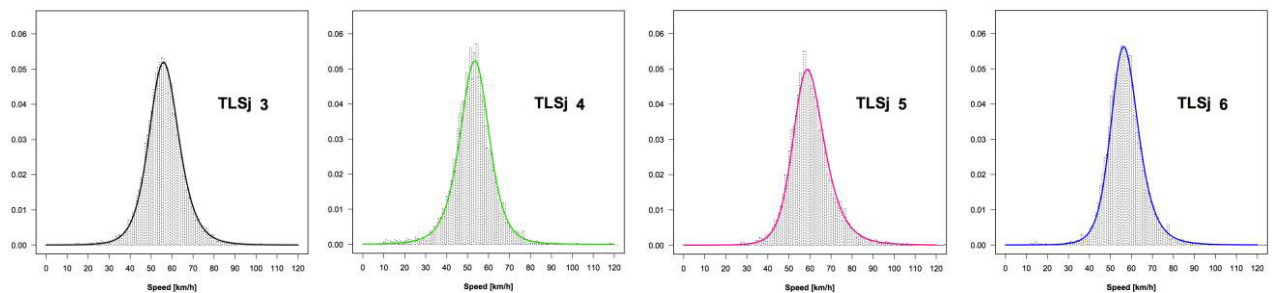


Fig. 5. Skew-T single models versus empirical distributions in different traffic conditions.

3. FR effects on VS

In view of the findings of previous section, we focus attention on FR effects on the shape of the ST models. Estimated ST parameters are shown in Table 7.

Table 7. Skew-T model parameters for various traffic conditions.

	ξ (location parameter)	σ^2 (scale parameter)	λ (skewness parameter)	ν (degrees of freedom)
TLS _j 1	52.56	12.42	1.53	8.35
TLS _j 2	53.87	6.68	0.215	4.42
TLS _j 3	54.11	7.58	0.361	4.99
TLS _j 4	54.45	7.3	-0.205	4.44
TLS _j 5	54.74	9.08	0.964	4.98
TLS _j 6	53.51	7.5	0.691	4.82

Our analysis focused on various combinations of FR_{pce} across directions (see Fig. 3), i.e.:

- Balanced. Balanced FR_{pce,1} and FR_{pce,2} (TLS_j 1, 3, 4, 5);

- Unbalanced 1. Same level of $FR_{pce,1}$ and different levels of $FR_{pce,2}$ (TLS_j 6-3): effect on VS of flow rate in the opposite direction;
- Unbalanced 2. Same level of $FR_{pce,2}$ and different levels of $FR_{pce,1}$ (TLS_j 5-6 and 3-2): effect on VS of flow rate in the analysed direction.

Balanced FR. Fig. 6 plots the models estimated in situations in which the values of FR_{pce} in the analysed and opposite directions are almost the same.

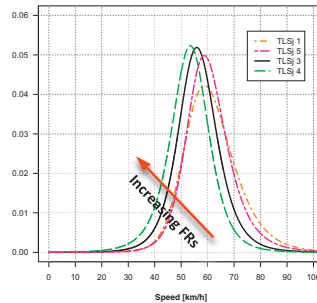


Fig. 6. Skew-T single models for balanced traffic conditions.

As the balanced $FR_{pce,1}$ and $FR_{pce,2}$ increase, the curves move to the left and the peaks rise: scale parameter σ^2 (Table 6) decreases from 12.42 km/h (TLS_j1) to 7.3 km/h (TLS_j4). This result is consistent with literature reports.

Unbalanced FR1. Fig. 8 shows the comparison between ST curves estimated in traffic conditions TLS_j 3 and 6. The effect of impedance produced by the opposite flow on the analysed one is clear-cut in the left branch of the models, and the skewness parameter does decrease from 0.691 to 0.361.

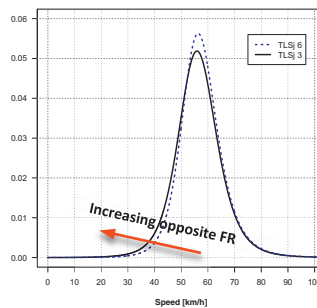


Fig. 7. Skew-T single models. Unbalanced traffic conditions 1.

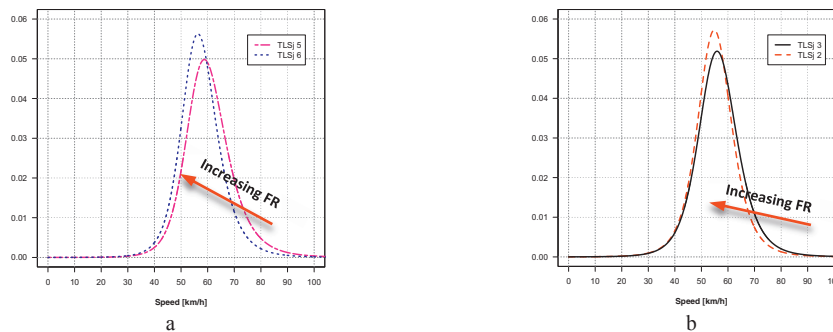


Fig. 8. Skew-T single models. Unbalanced traffic conditions 2.

Unbalanced FR2. With reference to TLS_j, 5 vs 6 and 3 vs 2, Figs. 8a and 8b plot ST single models. In both situations, as the FR increases in the analysed direction, the curves move to the left and become narrower. This tendency is confirmed by the reduction in the values of σ^2 , λ and v .

4. Summary and conclusions

This work presents a procedure for analysing vehicle speed (VS) distributions at cross-sections on two-way two-lane roads. Data (vehicle time headways, speeds and lengths) were collected at a cross-section along a perfectly straight road segment in flat terrain, with lane width 3.60 m and shoulder width 0.60 m.

The procedure was based on trendless analysis of time headways, in order to build trendless samples (TLS). The Passenger Car Equivalent (PCE) was estimated with a well-known approach, and used to estimate homogeneous flow rates in both lanes (FR_{pce1} , FR_{pce2}). Each TLS was classified according to the corresponding FR_{pce1} and FR_{pce2} .

Birch clustering methods were then used to find groups of similar TLSs in terms of FR_{pce} in both directions, and six groups were detected. Aggregate samples, called TLS_j, were created by joining the TLSs grouped in the same cluster, and were representative of the main traffic conditions of the analysed cross-section in the time period under analysis. In this way, large samples of homogeneous data were available for VS pdf fitting.

Three models were chosen to represent VS distribution: Normal (N), Skew-Normal (SN) and Skew-T (ST). In order to capture the multi-modality of speed distributions, a mixture modeling approach was used, and models with up to three components were analysed. The goodness-of-fit of the calibrated models was evaluated by three indicators: K-S distance, Pearson's chi-squared test, and the Bayesian Information Criterion (BIC).

In the case in question, regardless of traffic conditions, the Skew-T single model turned out to be the one which best fitted the observed phenomena: it represents the best trade-off between goodness-of-fit and model complexity.

The effect of flow rate in both directions (analysed and opposite) on VS distributions was also analysed. As expected, in all traffic conditions (balanced and unbalanced FRs across directions), the curves move to the left and become narrower as the FR increases. Examining both the estimated parameters and the shape of the curves, this effect seems to be less pronounced, where we measure the effect of the opposite flow rate.

Further analyses, such as transferability analysis of the models among samples and the application of the procedure to a wider set of cross-sections, are necessary in order to generalize the above results.

References

- Charest, L., Harrington, J., Salibian-Barrera, M., 2012. Birch: Dealing with very large datasets using BIRCH. R package version 1.2-3.
- Cox, D. R., Lewis, P. A. W., 1966. The Statistical Analysis of Series of Events. Methuen & Co Ltd., London.
- Dempster, A. P., Laird, N. M., Rubin, D. B., 1977. Maximum Likelihood from Incomplete Data via the EM Algorithm. Journal of the Royal Statistical Society Series B, 39, 1-38.
- Dey, P.P., Chandra, S., Gangopadhyaya, S., 2006. Speed distribution curves under mixed traffic conditions. Journal of Transportation Engineering, 132(5), 475-481.
- Gerlough, D. L., Huber, M. J., 1976. Special Report 165: Traffic Flow Theory. TRB, National Research Council, Washington, D.C..
- Ha, D.H., Aron, M., Cohen, S., 2011. Variations in parameters of time headway models according to macroscopic variables, fundamental diagram, and exogenous effects. Transportation Research Record: Journal of the Transportation Research Board, 2260, Transportation Research Board of the National Academies, Washington, D.C., 102-112.
- Jun, J., 2010. Understanding the Variability of Speed Distributions Under Mixed Traffic Conditions Caused by Holiday Traffic. Transportation Research Part C, Vol. 18, 4, 599-610.
- Luttinen, R. T., 1996. Statistical analysis of vehicle time headways. Publication 87, Helsinki University of Technology, Transp. Eng., Otaniemi.
- Luttinen, R.T., 2001. Traffic flow on two-lane highways - an overview. TL Research Report 1/2001, TL Consulting Engineers, Ltd., Lahti.
- Prates, M., Lachos, V., Cabral, C., 2013. Mixmsn: Fitting finite mixture of scale mixture of skew-normal distributions. R package version 1.0-8.
- Rossi, R., Gastaldi, M., 2012. An Empirical Analysis of Vehicle Time Headways on Rural Two-lane Two-way Roads. Procedia Social and Behavioral Sciences, 54, 865-874, doi: 10.1016/j.sbspro.2012.09.802.
- Rossi, R., Gastaldi, M., Pascucci, F., 2014. Gamma-GQM Time Headway Model: Endogenous Effects in Rural Two-lane Two-way Roads. Procedia Social and Behavioral Sciences, 859-868, doi: 10.1016/j.sbspro.2014.01.120.
- Zhang, T., Ramakrishnan, R., Livny, M., 1996. BIRCH: An efficient data clustering method for very large databases. SIGMOD.
- Zou, Y., Zhang Y., 2011. Use of Skew-normal and Skew-t distributions for mixture modelling of freeway speed data. Transportation Research Record: Journal of the Transportation Research Board, 2260, Transportation Research Board of the National Academies, Washington, D.C., 67-75.
- Zou, Y., Zhang Y., Zhu X., 2012. Constructing a bivariate distribution for freeway speed and headway data. Transportmetrica, 1-18.